

Deep learning applications in automatic segmentation and reconstruction in CT-based cervix brachytherapy

Hai Hu, MSc^{1,2}, Qiang Yang, PhD^{1,2}, Jie Li, MSc², Pei Wang, MSc², Bin Tang, PhD², Xianliang Wang, PhD², Jinyi Lang, MD²

¹Applied Nuclear Technology in Geosciences Key Laboratory of Sichuan Province, Chengdu University of Technology, Chengdu, China,

²Sichuan Cancer Hospital and Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Radiation Oncology Key Laboratory of Sichuan Province, Chengdu, China

The research was conducted in Department of Radiation Oncology, Sichuan Cancer Hospital and Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Radiation Oncology Key Laboratory of Sichuan Province, No. 55, the 4th Section, Renmin South Road, Chengdu, China.

Abstract

Purpose: Motivated by recent advances in deep learning, the purpose of this study was to investigate a deep learning method in automatic segment and reconstruct applicators in computed tomography (CT) images for cervix brachytherapy treatment planning.

Material and methods: U-Net model was developed for applicator segmentation in CT images. Sixty cervical cancer patients with Fletcher applicator were divided into training data and validation data according to ratio of 50 : 10, and another 10 patients with Fletcher applicator were employed to test the model. Dice similarity coefficient (DSC) and 95th percentile Hausdorff distance (HD95) were used to evaluate the model. Segmented applicator coordinates were calculated and applied into RT structure file. Tip error and shaft error of applicators were evaluated. Dosimetric differences between manual reconstruction and deep learning-based reconstruction were compared.

Results: The averaged overall 10 test patients' DSC, HD95, and reconstruction time were 0.89, 1.66 mm, and 17.12 s, respectively. The average tip error was 0.80 mm, and the average shaft error was less than 0.50 mm. The dosimetric differences between manual reconstruction and automatic reconstruction were 0.29% for high-risk clinical target volume (HR-CTV) $D_{90\%}$, and less than 2.64% for organs at risk D_{2cc} at a scenario of doubled maximum shaft error.

Conclusions: We proposed a deep learning-based reconstruction method to localize Fletcher applicator in three-dimensional CT images. The achieved accuracy and efficiency confirmed our method as clinically attractive. It paves the way for the automation of brachytherapy treatment planning.

J Contemp Brachytherapy 2021; 13, 3: 325-330

DOI: <https://doi.org/10.5114/jcb.2021.106118>

Key words: deep learning, applicator segmentation, brachytherapy, cervical cancer, dosimetric comparison.

Purpose

Cervical cancer is a common malignant tumor in women [1]. External beam radiotherapy combined with brachytherapy is the standard radiotherapy treatment for cervical cancer. At present, image-based three-dimensional (3D) brachytherapy has become the standard procedure. Applicator reconstruction is a critical step in treatment planning [2]. At this stage, the applicator reconstruction is performed manually by the planner.

Automatic planning is an important research topic in radiotherapy [3]. Automatic, accurate, and rapid appli-

cator reconstruction needs to be resolved for automatic planning in brachytherapy [4]. Before deep learning, researchers usually used threshold-based method to segment the applicator; however, this method still requires planners to define some points manually in clinical practice [5]. In recent years, more studies on automatic applicator reconstruction have been conducted based on deep learning [6-12].

In this study, a deep learning model to automatically segment and reconstruct the applicator was built. The dosimetric differences were compared between manual reconstruction and automatic reconstruction.

Address for correspondence: Xianliang Wang, PhD, Sichuan Cancer Hospital and Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Radiation Oncology Key Laboratory of Sichuan Province, No. 55, the 4th Section, Renmin South Road, Chengdu, China, phone: +86 138-804-097-33, ✉ e-mail: wangliu8687@163.com

Received: 11.12.2020

Accepted: 06.04.2021

Published: 13.05.2021

Material and methods

Segmentation model

The network structure is illustrated in Figure 1. The model was based on a U-Net structure and consisted of 10 components [13]. The model input included two-dimensional (2D) computed tomography (CT) images; the first 5 layers were under sampling layers, each layer contained two convolution operations and max pooling operation. Layers 6 to 9 were up sampling layers. After the transposed convolution operation, each layer merged the low-level information with the high-level information through a skip connection and then, performed two convolution operations. The convolution kernel was 3×3 , the transposed convolution kernel and the max pooling kernel were 2×2 , the step was both 1, and the activation function was Relu. The 10th layer integrated the cross-channel features through a 1×1 convolution layer, the activation function was sigmoid, and finally derived the desired 256×256 2D mask image.

Data annotation

We retrospectively studied 70 patients who completed CT-based 3D brachytherapy. All patients used a CT/magnetic resonance imaging (MRI) Fletcher applicator (Elekta part # 189.730). The resolution of CT images was $1 \text{ mm} \times 1 \text{ mm}$, and the slice thickness was 3 mm. The number of CT slices was 69 to 100 (average, 87). The 70 patients were divided into training data, validation data, and test data according to the ratio of 50 : 10 : 10. The annotating of applicator was performed by an experienced physicist using Oncentra (Elekta AB, Stockholm, Sweden, version 4.3) treatment planning system. The tandem diameter of CT/MRI Fletcher applicator was 4 mm, and the inner lumen of tandem was about 2 mm, so the applicator from the middle of each channel was annotated and depicted in a circle with a 2 mm radius. The number of CT slices containing mask images (ground truth) was from 38 to 71, and the average number of slices for training data and validation data was 58 and 56, respectively.

Data pre-process

We performed historical equalization on the CT images in the training data and validation data to raise

the applicator characteristics. For cervical cancer patients, the applicator is usually located in the middle area of CT image. In order to reduce the training data and validation data size, we shortened the CT images to a small region that included the Fletcher applicator (256×256 pixels). The shorten center was the geometric center of the CT image. We normalized all CT images and ground truths, so that all data were between 0 and 1. Deep learning model generally requires a large number of training data to learn effectively and prevent under-fitting or over-fitting. Therefore, we used ImageDataGenerator interface of Keras to augment the training data. More images were generated by rotating, enlarging, scaling, and shifting the image in other directions. The final training data and validation data were $256 \times 256 \times 116 \times 50$ and $256 \times 256 \times 56 \times 10$, respectively.

Training and segmentation

The training and segmentation were completed on an Intel Core i7-7700HQ CPU @ 2.80GHz, GPU NVIDIA GeForce GTX 1050ti, 8GB RAM personal computer, and the process is illustrated in Figure 2. In the training stage, the pre-processed training data and validation data were placed into the U-Net model. The batch size was 8, the epoch was 200, and the initial learning rate was 0.0001. The learning rate was dynamically adjusted by monitoring the learning process, and early stopping was adopted to avoid over-fitting of the model. We selected an Adam (adaptive moment estimation) as the optimizer and a dice loss as the loss function in the training stage [14]. The dice similarity coefficient (DSC) and dice loss were defined as follows:

$$DSC = \frac{(2|A \cap B| + \lambda)}{(|A| + |B| + \lambda)} \quad (1)$$

where A is the prediction mask image, B is the ground truth, and λ is the Laplace smoothing factor (usually 1), which could reduce over-fitting and avoid the denominator, and is 0.

$$L_s = 1 - DSC \quad (2)$$

After the training, the data of the test set were put into the model for segmentation, and the segmentation result of the corresponding patient was obtained and evaluated.

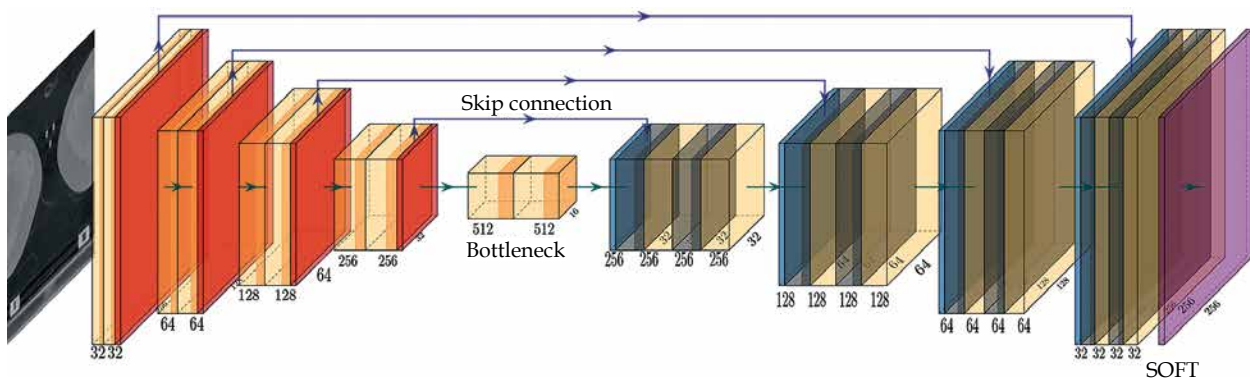


Fig. 1. U-Net structure for applicator segmentation

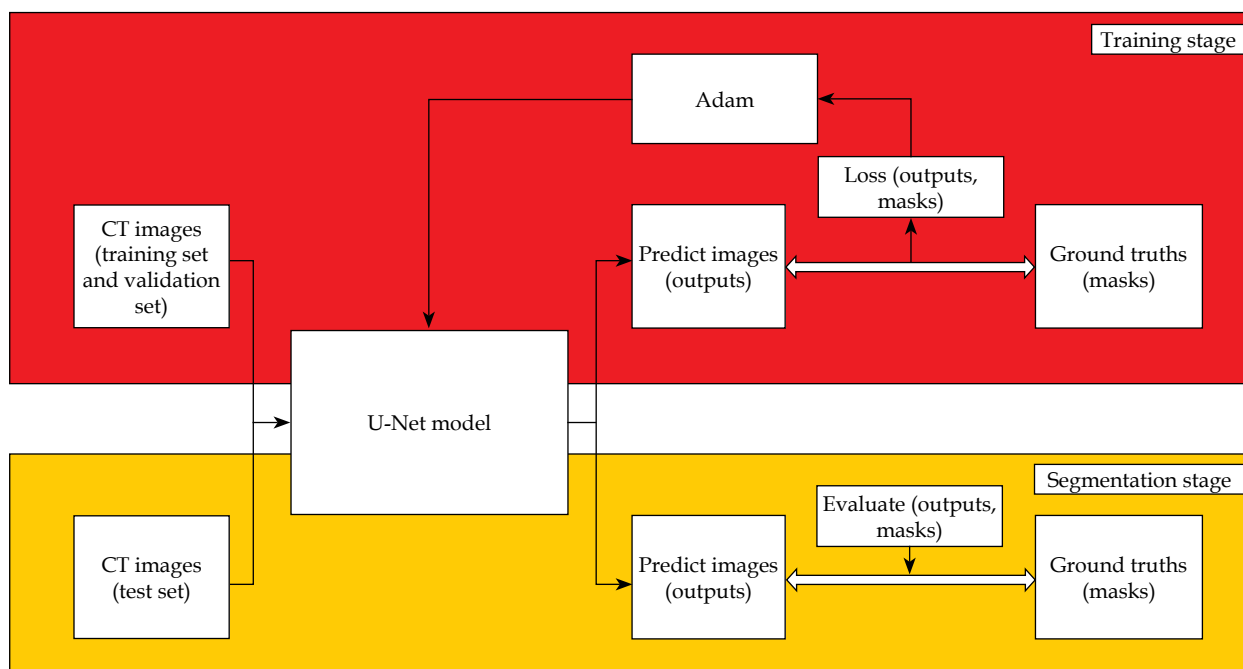


Fig. 2. Process of training and segmentation on the U-Net

Evaluation of segmentation results

We used two parameters to evaluate segmentation results [15, 16]. The first one was DSC. The DSC was used to measure the similarity of the two segmentation samples (manual and automatic). The range of DSC was 0-1. The larger the DSC, the better the segment. In the test process, the mean and standard deviation of DSC were calculated for all slices.

The second parameter was Hausdorff distance (HD). The HD was defined as:

$$HD(A, B) = \max(D(A, B), D(B, A)) \tag{3}$$

$$D(A, B) = \max_{a \in A} \min_{b \in B} ||a-b|| \tag{4}$$

where A is the prediction image surface, B is the ground truth surface, and a and b are the points on the surfaces of A and B, respectively. In order to eliminate the influence of outliers between the predicted image and the ground truth, the 95 percentile HD (HD95) was calculated. The unit of HD95 was mm. The smaller the HD95, the better the segmentation.

Automatic applicator reconstruction

We applied the test data into the trained model. The segmented applicator contour was composed by multiple points. We used a clustering method to create applicator contours as showed in reference [12]. For each channel, the average coordinate value of all points in one slice was calculated to obtain the trajectory of the channel central path. A polynomial curve fitting method was used in the reconstruction to reduce the systematic error. Then, the trajectory was written into the RT structure file, and the process of automatic reconstruction was completed.

Evaluation of reconstruction results

For each patient, we used the tip error and the shaft error to evaluate the reconstruction results [17]. The definitions of tip error and shaft error were as follows:

$$E_{Tip} = \frac{1}{N} \sum_{i=1}^N |Pred_i - Gt_i| \tag{5}$$

where N is the total channel number (3 in this study), $Pred_i$ is the predicted length of the i-th channel, and Gt_i is the annotation length of the i-th channel.

$$E_{Shaft} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^n ||Pred(x, y), Gt(x, y)|| \tag{6}$$

where M is the slices number, $Pred(x, y)$ are the predicted coordinates of the i-th slice, and $Gt(x, y)$ are the annotation coordinates of the i-th slice.

Dosimetric comparison

Dose volume histogram (DVH) parameters were used to evaluate the dosimetric difference between the automatic reconstruction and manual reconstruction. The DVH parameters were $D_{90\%}$ for high-risk clinical target volume (HR-CTV), and D_{2cc} for organs at risk (OARs). The OARs included bladder, rectum, sigmoid, and intestines [18].

Results

In the training stage, the loss converted to a lower level after 10 epochs. Because of the early stopping, the model finished training after 60 epochs. At the end of training stage, the loss of training data and validation data decreased to 0.10 and 0.11. The average DSC was 0.90 for the training data and 0.89 for the validation data.

The training results indicated that there was no over-fitting. The total training time was 8 hours for 60 epochs.

The segmentation and reconstruction results are presented in Table 1. In the average overall test data, the applicator segmented DSC was 0.89 and the HD95 was 1.66 mm. Compared with the manual reconstruction, the average tip error of the 10 cases was 0.80 mm, and the shaft errors were all within 0.50 mm. Tip error and shaft error of three channels were within a reliable range. Table 2 shows the breakdown time in this model. The average total time (including pre-processing, segmentation, and reconstruction) was 17.12 s. A reconstruction comparison of Fletcher applicator is illustrated in Figure 3.

In order to obtain a more conservative result, we chose 1 mm, which doubled the maximum shaft error (0.5 mm), to compare dosimetric differences. Table 3 pres-

ents the dosimetric data obtained by the two different reconstruction methods. Although we increased the shaft error, the dosimetric differences of HR-CTV $D_{90\%}$ were still less than 0.30%, and the maximum 2.64% for OARs D_{2cc} . These results confirmed that the accuracy of this model was acceptable [19].

Discussion

Applicator reconstruction is one of the most critical steps in brachytherapy treatment planning [2, 20]. Motivated by recent advances in deep learning, we investigated a deep learning method to automatically segment and reconstruct applicators in CT images for cervix brachytherapy treatment planning with the Fletcher applicator. Evaluation results proved its feasibility and reli-

Table 1. The results of applicator segmentation and reconstruction on the test cases

Test case	Segmentation			Reconstruction		
	DSC	HD95 (mm)	E_{Tip} (mm)	E_{Shaft} (mm)		
				Channel 1	Channel 2	Channel 3
1	0.88 ±0.10	2.07 ±5.28	1.00	0.38 ±0.37	0.49 ±0.33	0.32 ±0.26
2	0.90 ±0.07	0.97 ±0.83	0.00	0.29 ±0.23	0.50 ±0.32	0.30 ±0.22
3	0.90 ±0.09	1.40 ±2.44	1.00	0.35 ±0.37	0.32 ±0.34	0.30 ±0.17
4	0.89 ±0.09	1.26 ±2.19	0.00	0.30 ±0.25	0.35 ±0.46	0.33 ±0.20
5	0.88 ±0.12	1.56 ±3.69	1.00	0.30 ±0.28	0.28 ±0.32	0.31 ±0.22
6	0.88 ±0.08	1.89 ±6.27	1.00	0.36 ±0.29	0.40 ±0.26	0.36 ±0.34
7	0.89 ±0.10	2.07 ±5.56	2.00	0.45 ±0.77	0.47 ±0.58	0.27 ±0.17
8	0.90 ±0.06	0.99 ±0.76	1.00	0.34 ±0.16	0.30 ±0.24	0.33 ±0.16
9	0.89 ±0.13	1.72 ±4.31	1.00	0.26 ±0.12	0.29 ±0.19	0.26 ±0.26
10	0.90 ±0.09	2.66 ±8.90	0.00	0.30 ±0.25	0.31 ±0.24	0.25 ±0.16
Mean	0.89 ±0.09	1.66 ±4.02	0.80	0.33 ±0.31	0.37 ±0.33	0.30 ±0.22

Channel 1, Channel 2, Channel 3 – three channels of Fletcher applicator

Table 2. Breakdown time (s)

Test case	Pre-processing time	Segmentation time	Reconstruct time	Total time
1	3.64	5.54	7.14	16.32
2	3.33	6.17	7.06	16.56
3	2.84	5.85	7.36	16.05
4	3.27	5.98	8.16	17.41
5	4.15	5.19	10.01	19.35
6	3.14	5.90	7.93	16.97
7	3.55	5.43	7.48	16.46
8	3.18	5.39	7.08	15.65
9	4.12	6.08	9.00	19.20
10	3.54	5.82	7.84	17.20
Mean	3.48	5.73	7.91	17.12

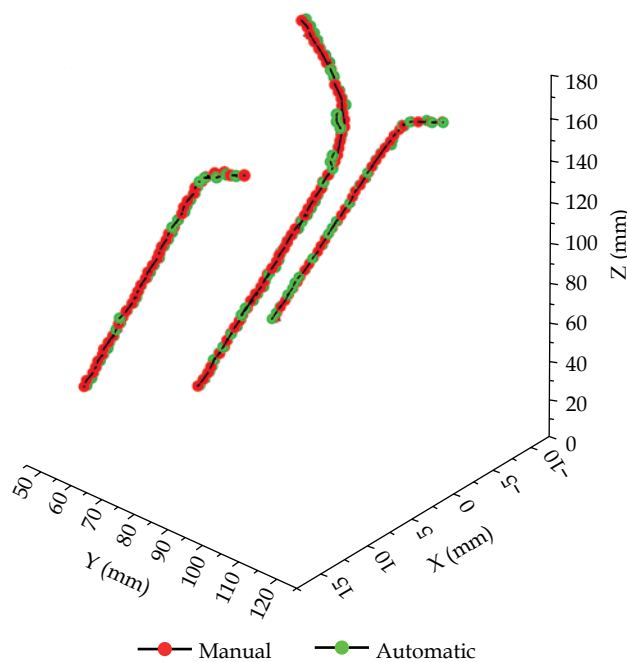


Fig. 3. Comparison of Fletcher applicator reconstructed by manual and automatic methods

ability. The model can quickly and accurately segment the applicator regions, and complete the reconstruction. For the commonly used applicator reconstruction, this model takes about 17.12 s from pre-processing to reconstruction, while an experienced physicist needs about 60 s, which increases the reconstruction efficiency by almost 4 times. The pre-processing, segmentation, and reconstruction time would be shorter if a high-performance computer was used.

Many scholars have conducted a series of research on the reconstruction of interstitial needles. Zhang *et al.* constructed an attention network and applied it to ultrasound-guided high-dose-rate prostate brachytherapy [6]. Wang *et al.* built two kinds of neural networks for segmentation of interstitial needles in ultrasound-guided prostate brachytherapy [7]. By constructing a 3D U-Net network, Zaffino *et al.* completed the reconstruction of interstitial needles in MRI-guided cervical cancer brachytherapy [17]. Moreover, Dai *et al.* developed a deeply supervised model by an attention-gated U-Net, incorporated with total variation regularization to detect multi-interstitial needles in MRI-guided prostate brachytherapy [8].

There are also studies about applicator segmentation. Hrinivich *et al.* studied an image model algorithm to reconstruct the applicator in cervical cancer guided by MRI. The average reconstruction accuracy of ring applicator and tandem applicator were 0.83 mm and 0.78 mm, respectively [9]. Based on the U-Net, Jung *et al.* proposed a deep learning-assisted applicators and interstitial needles digitization method for 3D CT image-based brachytherapy. In tandem and ovoid applicator digitization, DSC reached 0.93 and HD was less than 1 mm [10, 11]. Deufel *et al.* applied image thresholding and density-based clustering in applicator digitization. Their HDs were ≤ 1.0 mm, and the differences for HR-CTV $D_{90\%}$

Table 3. The results of dosimetric differences between manual and automatic reconstructions

	Parameters	Manual	Automatic	Differences
HR-CTV	$D_{90\%}$	600.42 \pm 0.82	598.70 \pm 3.55	0.29%
Rectum	D_{2cc}	339.29 \pm 44.88	334.96 \pm 41.69	1.27%
Bladder	D_{2cc}	392.48 \pm 43.06	402.83 \pm 43.87	2.64%
Sigmoid	D_{2cc}	273.81 \pm 93.05	274.77 \pm 93.01	0.35%
Intestines	D_{2cc}	350.78 \pm 65.65	353.64 \pm 65.41	0.82%

The unit of $D_{90\%}$ and D_{2cc} is cGy

$D_{95\%}$, and OARs D_{2cc} were less or equal to 1% [12]. In the present study, the DSC was 0.89, HD was 1.66 mm, the dosimetric differences for the target were less than 0.30%, and the maximum 2.64% for OARs D_{2cc} . Compared with previous studies, our results still have room for an improvement.

After the model trained with the Fletcher applicator, six patients with a vaginal CT/MRI applicator (Elekta part # 101.001) were also used to test this model. A vaginal applicator differs from the Fletcher applicator with a connection end. The average overall test data DSC, HD95, tip error, and shaft error were 0.84, 1.81 mm, 1.00 mm, and 0.31 mm, respectively. The dosimetric difference of HR-CTV $D_{90\%}$ was less than 0.51%, and the one of OARs D_{2cc} was less than 4.87%. The results of vaginal applicator were slightly worse than that of the Fletcher applicator, since the model was trained by Fletcher applicators; however, the differences of all evaluated parameters were less than 5%.

Automatic radiotherapy planning is a hot spot in current studies, and it is also a subject of interest of our research group. We have made efforts in this direction [21, 22]. These results prove that this model could be integrated into an automatic treatment planning system.

Our present study has some limitations. One is that, according to published research, the accuracy needs to be improved. Although the dosimetric differences between the two reconstruction methods were acceptable, we are still working on the ways to increase the segmentation and reconstruction accuracies. The other limitation is that only two applicator types were included in this study. Here, we chose the Fletcher applicator because it is one of the most commonly used applicators in our center. Even though there are many kinds of applicators in clinical practice, this model could not be used for other applicator types. However, we are convinced that reconstruction of other applicator types can be carried out quickly with the foundation of the present research. Another limitation of this work is that the CT slice thickness was large in this study (3 mm). Slice thickness is a source of the tip uncertainty; therefore, the large slice thickness could be a reason of the large tip error in this study.

Conclusions

In summary, applicator reconstruction is a critical process of treatment planning. We implemented a U-Net model for applicator segmentation and reconstruction in

CT-based cervix brachytherapy in this study. The DSC, HD95, tip error, and shaft error were used to evaluate this model. The results demonstrated that our model is clinically attractive. Therefore, this research paves the way for automatic treatment planning in brachytherapy.

Acknowledgements

This research was supported by National Key Research and Development Project (No. 2017YFC0113100), Chengdu Science and Technology Project (No. 2019-YF09-00095-SN), and Key Research and Development Project of Sichuan (No. 2019YFS0473).

Disclosure

The authors report no conflict of interest.

References

1. Bray F, Ferlay J, Soerjomataram I et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394-424.
2. Hellebust TP, Kirisits C, Berger D et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group: considerations and pitfalls in commissioning and applicator reconstruction in 3D image-based treatment planning of cervix cancer brachytherapy. *Radiother Oncol* 2010; 96: 153-160.
3. Fionda B, Boldrini L, D'Aviero A et al. Artificial intelligence (AI) and interventional radiotherapy (brachytherapy): state of art and future perspectives. *J Contemp Brachytherapy* 2020; 12: 497-500.
4. Shen C, Gonzalez Y, Klages P et al. Intelligent inverse treatment planning via deep reinforcement learning, a proof-of-principle study in high dose-rate brachytherapy for cervical cancer. *Phys Med Biol* 2019; 64: 115013.
5. Binny D, Lancaster CM, Kairn T et al. Investigating the use of image thresholding in brachytherapy catheter reconstruction. *Australas Phys Eng Sci Med* 2016; 39: 913-919.
6. Zhang Y, Lei Y, Qiu RLJ et al. Multi-needle Localization with Attention U-Net in US-guided HDR Prostate Brachytherapy. *Med Phys* 2020; 47: 2735-2745.
7. Wang F, Xing L, Bagshaw H et al. Deep learning applications in automatic needle segmentation in ultrasound-guided prostate brachytherapy. *Med Phys* 2020; 47: 3797-3805.
8. Dai X, Lei Y, Zhang Y et al. Automatic multi-catheter detection using deeply supervised convolutional neural network in MRI-guided HDR prostate brachytherapy. *Med Phys* 2020; 47: 4115-4124.
9. Hrinivich WT, Morcos M, Viswanathan A, Lee J. Automatic tandem and ring reconstruction using MRI for cervical cancer brachytherapy. *Med Phys* 2019; 46: 4324-4332.
10. Jung H, Gonzalez Y, Shen C et al. Deep-learning-assisted automatic digitization of applicators in 3D CT image-based high-dose-rate brachytherapy of gynecological cancer. *Brachytherapy* 2019; 18: 841-851.
11. Jung H, Shen C, Gonzalez Y et al. Deep-learning assisted automatic digitization of interstitial needles in 3D CT image based high dose-rate brachytherapy of gynecological cancer. *Phys Med Biol* 2019; 64: 215003.
12. Deufel CL, Tian S, Yan BB et al. Automated applicator digitization for high-dose-rate cervix brachytherapy using image thresholding and density-based clustering. *Brachytherapy* 2020; 19: 111-118.
13. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.
14. Kingma D, Ba J. Adam: a method for stochastic optimization. *Computer ence* 2014.
15. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* 2015; 15: 29.
16. Chang HH, Zhuang AH, Valentino DJ, Chu WC. Performance measure characterization for evaluating neuroimage segmentation algorithms. *Neuroimage* 2009; 47: 122-135.
17. Zaffino P, Pernelle G, Mastmeyer A et al. Fully automatic catheter segmentation in MRI with 3D convolutional neural networks: application to MRI-guided gynecologic brachytherapy. *Phys Med Biol* 2019; 64: 165008.
18. Pötter R, Haie-Meder C, Van Limbergen E et al. Recommendations from gynaecological (GYN) GEC ESTRO working group (II): concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy-3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology. *Radiother Oncol* 2006; 78: 67-77.
19. Nath R, Anderson LL, Meli JA et al. Code of practice for brachytherapy physics: Report of the AAPM Radiation Therapy Committee Task Group No. 56. *Med Phys* 1997; 24: 1557-1598.
20. Hansen J, Dunkerley D, Bradley K et al. Comparison of catheter reconstruction techniques for the lunar ovoid channels of the Venezia(TM) applicator. *J Contemp Brachytherapy* 2020; 12: 383-392.
21. Wang X, Wang P, Li C et al. An automated dose verification software for brachytherapy. *J Contemp Brachytherapy* 2018; 10: 478-482.
22. Wang X, Wang P, Tang B et al. An inverse dose optimization algorithm for three-dimensional brachytherapy. *Front Oncol* 2020; 10: 564580.